|  |  |
| --- | --- |
| Name | Boxplot of 1 feature for all jobs. |
| Category | Platform-wide analysis |
| Description | Box plot of 1 feature, for example job size (in number of processes) or job’s data size (number of bytes transferred per job), for all jobs. |
| Query | select nprocs from jobs\_all; |
| Possible Parameters | select clause: nprocs, total\_bytes, runtime, I/O time  where clause: start\_time within a period, uid, projid  filtering parameters: nprocs = 1, total\_bytes < 1024, |
| Chart Type | Box plot |
| Screenshot |  |

|  |  |
| --- | --- |
| Name | Boxplot of 1 feature for all applications. |
| Category | Platform-wide analysis |
| Description | Box plot of 1 feature, such as maximum I/O throughput that an application has ever achieved, for all applications. |
| Query | select real\_exe, max(nprocs) as max\_jobsize from jobs\_all group by exe\_name; |
| Possible Parameters | select clause: max(agg\_perf\_MB), max(nprocs), max(total\_bytes), avg(nprocs),  where clause: start\_time within a period, uid, projid |
| Chart Type | Box plot |
| Screenshot |  |

|  |  |
| --- | --- |
| Name | I/O throughput of jobs/ applications on the platform |
| Category | Platform-wide analysis |
| Description | Jobs’ I/O throughput for all jobs or Maximum I/O throughput of each application for all applications.  X-axis could be in absolute count or percentage. |
| Query | Q1: select agg\_perf\_MB from jobs\_all where agg\_perf\_MB is not null order by agg\_perf\_MB desc;  Q2: select @appid:=@appid+1 as appid, @appid/q3.numapp as app\_percentage, q2.\* from (select @appid:=0) r, (select real\_exe, max(agg\_perf\_MB) as max\_perf from jobs\_all where agg\_perf\_MB is not null group by real\_exe order by max\_perf desc) q2, (select count(distinct real\_exe) as numapp from jobs\_all) q3; |
| Possible Parameters | Select clause: agg\_perf\_MB,  where clause: start\_time within a period, uid, projid |
| Chart Type | Line chart, scatter |
| Screenshot | appmaxthruput |

|  |  |
| --- | --- |
| Name | Distribution of jobs size and jobs data size for all jobs. |
| Category | Platform-wide analysis |
| Description |  |
| Query | select total\_bytes, agg\_perf\_MB from jobs\_all where agg\_perf\_MB is not null;  @Amir:  select round(log10(total\_bytes), 1) as tb, round(log10(agg\_perf\_MB),1) as ap, count(\*) cc from jobs\_all where agg\_perf\_MB is not null group by tb, ap order by cc desc; |
| Possible Parameters | Select clause: total\_bytes, agg\_perf\_MB, nprocs  Where clause: |
| Chart Type | Scatter in heatmap. The color indicates number of jobs of a certain number of a certain (total\_bytes, perf). |
| Screenshot |  |
| R script | This is the script using R to create this graph: main idea: using stat\_binhex  ggplot(moverall, aes(log2(total\_bytes), log2(agg\_perf\_MB))) +  stat\_binhex(aes(fill = cut(..count.., c(0, 10, 100, 500, 1000, 5000,10000, 50000,100000), labels = c('1 - 10', '11 - 100','101 - 500', '501 - 1k', '1k1 - 5k','5k1 - 10k','10k1 - 50k','50k1-100k'))))  +  xlab("Number of bytes transferred") + ylab("I/O Throughput") + scale\_y\_continuous(breaks = seq(-20,20,10),labels=c("1 B/s", "1 KB/s", "1 MB/s", "1 GB/s","1TB/s")) +  scale\_x\_continuous (breaks = seq(0,50,10),labels=c("1 B", "1 KB", "1 MB", "1 GB","1 TB","1 PB"))  + geom\_hline(yintercept=log2(1024 \* 240),colour="darkgreen",size=1) +  geom\_hline(yintercept=log2(2560 ),colour="black",size=1) + geom\_hline(yintercept=log2(256),colour="purple",size=1)   + theme\_bw() + theme(axis.text=element\_text(size=14,color="black"),axis.title=element\_text(size=24), panel.grid.major = element\_line(colour = "grey40"), plot.title=element\_text(size=24))  + annotate("text", x = 15, y = 19, label = "System peak - 240 GB/s")+ annotate("text", x = 6, y = 12.5, label ="10 USB") + annotate("text", x = 6, y = 9, label ="1 USB") + ggtitle("Mira: Jobs I/O Throughput")  +  labs(fill = 'Jobs Count') +  scale\_fill\_manual(values = c("cadetblue1","dodgerblue","blue","green","yellow","darkorange","red",'darkorchid4')) |

|  |  |
| --- | --- |
| Name |  |
| Category | Platform-wide analysis |
| Description |  |
| Query |  |
| Possible Parameters |  |
| Chart Type |  |
| Screenshot |  |

* 1. Distribution of jobs size and jobs data size for all jobs.

The hard part for this one is create this heatmap-type of graph.

1. Category: Application-specific analysis