CLAIMS

What is claimed:

1. A method for performing backup of electronic data, the method comprising:

sampling metadata of files on a computer at a predetermined time interval;

deriving a model of change of the metadata over time, wherein the model of change is a snapshot of dynamically sampled attributes of the files and reflects a state of the computer over the predetermined time interval;

performing an analysis of the model of change by determining features of the files, the features comprising file usage dynamics of both the computer and a user associated with the files, wherein the analysis further comprises predicting a change in file size of the files and a change in capacity of the computer subsequent to the predetermined time interval;

generating an urgency model based on the metadata further indicating data access frequency and periodicity;

deriving an urgency of each of the files based on the urgency model, the urgency indicating (1) how soon a file must be retrieved from backup and (2) conditions on whether to back up the file in low latency storage or high latency storage;

selecting a backup plan for the files according to the urgency of each of the files, the change in file size, and the change in capacity of the computer; and

backing up files to a storage server according to the selected backup plan.

2. The method of claim 1, further comprising:

predicting the file usage dynamics of the computer and the user based on the model of change.

3. The method of claim 1, further comprising one or more of:

predicting urgency of new files on the computer based on the urgency model; and

predicting attributes of new files on the computer based on the urgency model.

4. The method of claim 3, wherein the attributes comprise one or more of data type and data size.

5. The method of claim 3, further comprising:

predicting a growth of a particular file type.

6. The method of claim 1, further comprising one or more of:

predicting, using the model of change and the urgency model, which kinds of files are subjected to particular file operations more often than other kinds of files;

predicting which kinds of files grow in a monotonic manner;

predicting which kinds of files are periodically or un-periodically accessed; and

predicting file operations that cause change in system storage capacity or backup, the change being greater than a predetermined threshold.

7. The method of claim 1, further comprising;

deriving derivative features from attributes of the metadata; and

predicting the file usage dynamics based on a combination of the derivative features and the attributes of the metadata.

8. The method of claim 7, wherein deriving derivative features further comprises:

analyzing values in the attributes of the metadata;

encoding the values that were analyzed into encoded values in a particular format;

bounding the encoded values according to the format; and

deriving derivative the features from the attributes in the metadata according to the encoded values.

9. A system for performing backup of electronic data, comprising:

a hardware processor configured to:

sample metadata of files on a computer at a predetermined time interval;

derive a model of change of the metadata over time, wherein the model of change is a snapshot of dynamically sampled attributes of the files and reflects a state of the computer over the predetermined time interval;

perform an analysis of the model of change by determining features of the files, the features comprising file usage dynamics of both the computer and a user associated with the files, wherein the analysis further comprises predicting a change in file size of the files and a change in capacity of the computer subsequent to the predetermined time interval;

generate an urgency model based on the metadata further indicating data access frequency and periodicity;

derive an urgency of each of the files based on the urgency model,

the urgency indicating (1) how soon a file must be retrieved from backup and (2) conditions on whether to back up the file in low latency storage or high latency storage;

select a backup plan for the files according to the urgency of each of the files, the change in file size, and the change in capacity of the computer; and

back up files to a storage server according to the selected backup plan.

10. The system of claim 9, wherein the hardware processor is further configured to:

predict the file usage dynamics of the computer and the user based on the model of change.

11. The system of claim 9, wherein the hardware processor is further configured to:

predict urgency of new files on the computer based on the urgency model; and

predict attributes of new files on the computer based on the urgency model.

12. The system of claim 11, wherein the attributes comprise one or more of data type and data size.

13. The system of claim 11, wherein the hardware processor is further configured to:

predict a growth of a particular file type.

14. The system of claim 9, wherein the hardware processor is further configured to perform one or more of:

predict, using the model of change and the urgency model, which kinds of files are subjected to particular file operations more often than other kinds of files;

predict which kinds of files grow in a monotonic manner;

predict which kinds of files are periodically or un-periodically accessed; and

predict file operations that cause change in system storage capacity or backup, the change being greater than a predetermined threshold.

15. The system of claim 9, wherein the hardware processor is further configured to;

derive derivative features from attributes of the metadata; and

predict the file usage dynamics based on a combination of the derivative features and the attributes of the metadata.

16. The system of claim 15, wherein the hardware processor is further configured to derive the derivative features by:

analyzing values in the attributes of the metadata;

encoding the values that were analyzed into encoded values in a particular format;

bounding the encoded values according to the format; and

deriving derivative the features from the attributes in the metadata according to the encoded values.

17. A non-transitory computer-readable medium storing thereon instructions that when executed perform a method for backup of electronic data, the instructions comprising:

sampling metadata of files on a computer at a predetermined time interval;

deriving a model of change of the metadata over time, wherein the model of change is a snapshot of dynamically sampled attributes of the files and reflects a state of the computer over the predetermined time interval;

performing an analysis of the model of change by determining features of the files, the features comprising file usage dynamics of both the computer and a user associated with the files, wherein the analysis further comprises predicting a change in file size of the files and a change in capacity of the computer subsequent to the predetermined time interval;

generating an urgency model based on the metadata further indicating data access frequency and periodicity;

deriving an urgency of each of the files based on the urgency model, the urgency indicating (1) how soon a file must be retrieved from backup and (2) conditions on whether to back up the file in low latency storage or high latency storage;

selecting a backup plan for the files according to the urgency of each of the files, the change in file size, and the change in capacity of the computer; and

backing up files to a storage server according to the selected backup plan.

18. The medium of claim 17, the instructions further comprising:

predicting the file usage dynamics of the computer and the user based on the model of change.